ONCOMINE: A Cancer Microarray Database and Web-Based Data-Mining Platform

Rhodes, Daniel R.*[*1], *Yu, Jianjun*[1,2], *K., Shanker*[3], *Deshpande, Nandan*[3], *Varambally, Radhika*[1], *Ghosh, Debashis*[5], *Barrette, Terrence*[1], *Pandey, Akhilesh*[4], *Chinnaiyan, Arul M.*[1,6,7]

[1]Departments of Pathology, [2]Bioinformatics, [5]Biostatistics, [6]Urology, and [7]Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI, USA; [3]Institute of Bioinformatics, Bangalore, India; [4]McKusick-Nathans Institute of Genetic Medicine and Department of Biological Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD, USA

DNA microarray technology has led to an explosion of oncogenomic analyses, generating a wealth of data and uncovering the complex gene expression patterns of cancer. Unfortunately, due to the lack of a unifying bioinformatics resource, the majority of this data sits stagnant and disjointed following publication, massively under-utilized by the cancer research community. Here, we present *ONCOMINE*, a cancer microarray database and web-based data-mining platform aimed at unifying data and analysis, and facilitating discovery from genome-wide expression analyses. To date, *ONCOMINE* contains 65 cancer gene expression datasets, comprising nearly 50 million gene expression measurements form more than 4,700 microarray experiments. From this data, more than 100 differential expression analyses have been performed, identifying the genes most over- and under-expressed in most types of cancer relative to respective normal tissue as well as in a variety of cancer subtypes, including many defined by clinical and pathological criteria. The data and differential expression analyses can be queried and visualized with one of three query modules, GENE, STUDY, and META. In the GENE module, the expression of a gene of interest can be easily assessed across all datasets, allowing the user to quickly identify in which cancer types or subtypes the gene of interest is differentially expressed. In the STUDY module, the genes most over- or under-expressed in a selected cancer type or subtype can be assessed and visualized. Finally, in the META module, the genes most commonly differentially expressed in multiple similar datasets can be explored. To rapidly interpret a gene's potential role in cancer, *ONCOMINE* also includes a gene annotation repository, centralizing information from many genome resources including Unigene, SwissProt, and LocusLink. Further, by integrating expression analyses with gene ontology annotations and a therapeutic target database, *ONCOMINE* allows the user to focus on clinically important genes that are differentially expressed in cancer, including: secreted, kinase, membrane-bound, and known drug target, facilitating the discovery of potential cancer markers and therapeutic targets. In summary, *ONCOMINE* provides a unified infrastructure for cancer gene expression analysis and will likely promote the increased utilization of and hypothesis generation from cancer microarray data, ultimately leading to an improved understanding of cancer and the development of novel diagnostic and therapeutic strategies.